

Pandas Basics

November 30, 2024

1 Pandas Basics

```
[64]: import numpy as np
```

```
[3]: data = np.genfromtxt('example_data.csv', delimiter=';', names=True, dtype=None,
    ↪encoding='UTF')
```

```
[4]: data
```

```
[4]: array([('2018-10-13 11:10:23.560', '262km NW of Ozernovskiy, Russia', 'mww',
    6.7, 'green', 1),
    ('2018-10-13 04:34:15.580', '25km E of Bitung, Indonesia', 'mww', 5.2,
    'green', 0),
    ('2018-10-13 00:13:46.220', '42km WNW of Sola, Vanuatu', 'mww', 5.7,
    'green', 0),
    ('2018-10-12 21:09:49.240', '13km E of Nueva Concepcion, Guatemala',
    'mww', 5.7, 'green', 0),
    ('2018-10-12 02:52:03.620', '128km SE of Kimbe, Papua New Guinea', 'mww',
    5.6, 'green', 1)],
    dtype=[('time', '<U23'), ('place', '<U37'), ('magType', '<U3'), ('mag',
    '<f8'), ('alert', '<U5'), ('tsunami', '<i8')])
```

```
[5]: data.shape
```

```
[5]: (5,)
```

```
[6]: data.dtype
```

```
[6]: dtype([('time', '<U23'), ('place', '<U37'), ('magType', '<U3'), ('mag', '<f8'),
    ('alert', '<U5'), ('tsunami', '<i8')])
```

```
[23]: %%timeit
    max([row[3] for row in data])
```

1.77 s ± 13.8 ns per loop (mean ± std. dev. of 7 runs, 1,000,000 loops each)

```
[8]: array_dict = { col: np.array([row[i] for row in data]) for i, col in
    ↪enumerate(data.dtype.names)}
```

```
[9]: array_dict
```

```
[9]: {'time': array(['2018-10-13 11:10:23.560', '2018-10-13 04:34:15.580',  
                  '2018-10-13 00:13:46.220', '2018-10-12 21:09:49.240',  
                  '2018-10-12 02:52:03.620'], dtype='<U23'),  
      'place': array(['262km NW of Ozernovskiy, Russia', '25km E of Bitung,  
Indonesia',  
                    '42km WNW of Sola, Vanuatu',  
                    '13km E of Nueva Concepcion, Guatemala',  
                    '128km SE of Kimbe, Papua New Guinea'], dtype='<U37'),  
      'magType': array(['mww', 'mww', 'mww', 'mww', 'mww'], dtype='<U3'),  
      'mag': array([6.7, 5.2, 5.7, 5.7, 5.6]),  
      'alert': array(['green', 'green', 'green', 'green', 'green'], dtype='<U5'),  
      'tsunami': array([1, 0, 0, 0, 1])}
```

```
[10]: array_dict['mag'].max()
```

```
[10]: 6.7
```

```
[11]: np.array([value[array_dict['mag'].argmax()] for key, value in array_dict.  
               ↪items()])
```

```
[11]: array(['2018-10-13 11:10:23.560', '262km NW of Ozernovskiy, Russia',  
            'mww', '6.7', 'green', '1'], dtype='<U32')
```

```
[12]: import pandas as pd
```

```
[13]: place = pd.Series(array_dict['place'], name='place')  
place
```

```
[13]: 0      262km NW of Ozernovskiy, Russia  
1      25km E of Bitung, Indonesia  
2      42km WNW of Sola, Vanuatu  
3      13km E of Nueva Concepcion, Guatemala  
4      128km SE of Kimbe, Papua New Guinea  
Name: place, dtype: object
```

```
[14]: place_index = place.index  
place_index
```

```
[14]: RangeIndex(start=0, stop=5, step=1)
```

```
[15]: place_index.values
```

```
[15]: array([0, 1, 2, 3, 4])
```

```
[16]: numbers = np.linspace(0,10, num=5)
      numbers
```

```
[16]: array([ 0. ,  2.5,  5. ,  7.5, 10. ])
```

```
[17]: x= pd.Series(numbers)
      x
```

```
[17]: 0    0.0
      1    2.5
      2    5.0
      3    7.5
      4   10.0
      dtype: float64
```

```
[18]: y = pd.Series(numbers, index=pd.Index([1,2,3,4,5]))
      y
```

```
[18]: 1    0.0
      2    2.5
      3    5.0
      4    7.5
      5   10.0
      dtype: float64
```

```
[19]: x + y
```

```
[19]: 0    NaN
      1    2.5
      2    7.5
      3   12.5
      4   17.5
      5    NaN
      dtype: float64
```

```
[20]: df = pd.DataFrame(array_dict)
      df
```

```
[20]:
```

| | time | place | magType | \ |
|---|-------------------------|---------------------------------------|---------|---|
| 0 | 2018-10-13 11:10:23.560 | 262km NW of Ozernovskiy, Russia | mww | |
| 1 | 2018-10-13 04:34:15.580 | 25km E of Bitung, Indonesia | mww | |
| 2 | 2018-10-13 00:13:46.220 | 42km WNW of Sola, Vanuatu | mww | |
| 3 | 2018-10-12 21:09:49.240 | 13km E of Nueva Concepcion, Guatemala | mww | |
| 4 | 2018-10-12 02:52:03.620 | 128km SE of Kimbe, Papua New Guinea | mww | |

| | mag | alert | tsunami |
|---|-----|-------|---------|
| 0 | 6.7 | green | 1 |

```

1  5.2  green      0
2  5.7  green      0
3  5.7  green      0
4  5.6  green      1

```

```
[22]: df.dtypes
```

```

[22]: time      object
      place      object
      magType    object
      mag        float64
      alert      object
      tsunami    int64
      dtype: object

```

```
[24]: df.values
```

```

[24]: array([[ '2018-10-13 11:10:23.560', '262km NW of Ozernovskiy, Russia',
               'mww', 6.7, 'green', 1],
             [ '2018-10-13 04:34:15.580', '25km E of Bitung, Indonesia', 'mww',
               5.2, 'green', 0],
             [ '2018-10-13 00:13:46.220', '42km WNW of Sola, Vanuatu', 'mww',
               5.7, 'green', 0],
             [ '2018-10-12 21:09:49.240',
               '13km E of Nueva Concepcion, Guatemala', 'mww', 5.7, 'green', 0],
             [ '2018-10-12 02:52:03.620', '128km SE of Kimbe, Papua New Guinea',
               'mww', 5.6, 'green', 1]], dtype=object)

```

```
[25]: df.columns
```

```
[25]: Index(['time', 'place', 'magType', 'mag', 'alert', 'tsunami'], dtype='object')
```

2 Arithmatic Operations on DataFrames

```
[26]: df + df
```

```

[26]:                                     time \
0  2018-10-13 11:10:23.5602018-10-13 11:10:23.560
1  2018-10-13 04:34:15.5802018-10-13 04:34:15.580
2  2018-10-13 00:13:46.2202018-10-13 00:13:46.220
3  2018-10-12 21:09:49.2402018-10-12 21:09:49.240
4  2018-10-12 02:52:03.6202018-10-12 02:52:03.620

                                     place magType  mag \
0  262km NW of Ozernovskiy, Russia262km NW of Oze... mwwmww  13.4
1  25km E of Bitung, Indonesia25km E of Bitung, I... mwwmww  10.4

```

| | | | | |
|---|---------------------------------------|--------------------------|--------|------|
| 2 | 42km WNW of Sola, Vanuatu | 42km WNW of Sola, Van... | mwwmww | 11.4 |
| 3 | 13km E of Nueva Concepcion, Guatemala | 13km E of... | mwwmww | 11.4 |
| 4 | 128km SE of Kimbe, Papua New Guinea | 128km SE of... | mwwmww | 11.2 |

| | alert | tsunami |
|---|------------|---------|
| 0 | greengreen | 2 |
| 1 | greengreen | 0 |
| 2 | greengreen | 0 |
| 3 | greengreen | 0 |
| 4 | greengreen | 2 |

```
[29]: import datetime as dt
import numpy as np
import pandas as pd
np.random.seed(0) # set seed so result is reproducible

pd.DataFrame(
    {
        'random': np.random.rand(5), 'text': ['hot', 'warm', 'cool', 'cold', None],
        'truth': [np.random.choice([True, False])
                  for _ in range(5)]
    },
    index=pd.date_range(end=dt.date(2019, 4, 21),freq='1D', periods=5,
        name='date')
)
```

```
[29]:          random  text  truth
date
2019-04-17  0.548814   hot  False
2019-04-18  0.715189  warm   True
2019-04-19  0.602763  cool   True
2019-04-20  0.544883  cold  False
2019-04-21  0.423655  None   True
```

```
[30]: pd.DataFrame([
        {'mag': 5.2, 'place': 'California'},
        {'mag': 1.2, 'place': 'Alaska'},
        {'mag': 0.2, 'place': 'California'},
    ])
```

```
[30]:   mag    place
0  5.2  California
1  1.2    Alaska
2  0.2  California
```

```
[32]: list_of_tuples = [(n, n**2, n**3) for n in range(5)]
list_of_tuples
```

```
[32]: [(0, 0, 0), (1, 1, 1), (2, 4, 8), (3, 9, 27), (4, 16, 64)]
```

```
[33]: pd.DataFrame(list_of_tuples, columns=['n', 'n_squared', 'n_cubed'])
```

```
[33]:
```

| | n | n_squared | n_cubed |
|---|---|-----------|---------|
| 0 | 0 | 0 | 0 |
| 1 | 1 | 1 | 1 |
| 2 | 2 | 4 | 8 |
| 3 | 3 | 9 | 27 |
| 4 | 4 | 16 | 64 |

```
[34]: pd.DataFrame(
    np.array([[0, 0, 0],[1, 1, 1],[2, 4, 8],[3, 9, 27],[4, 16, 64] ]),
    columns=['n', 'n_squared', 'n_cubed'])
```

```
[34]:
```

| | n | n_squared | n_cubed |
|---|---|-----------|---------|
| 0 | 0 | 0 | 0 |
| 1 | 1 | 1 | 1 |
| 2 | 2 | 4 | 8 |
| 3 | 3 | 9 | 27 |
| 4 | 4 | 16 | 64 |

```
[36]: !wc -l earthquakes.csv
```

```
9333 earthquakes.csv
```

```
[37]: !ls -lh | grep -i earth
```

```
-rw-rw-r--  1 test test 3.4M Nov 29 17:06 earthquakes.csv
```

```
[39]: files = !ls -lh data
[file for file in files if 'earthquake' in file]
```

```
[39]: []
```

```
[40]: !head -n 2 earthquakes.csv
```

```
alert,cdi,code,detail,dmin,felt,gap,ids,mag,magType,mmi,net,nst,place,rms,sig,so
urces,status,time,title,tsunami,type,types,tz,updated,url
,,37389218,https://earthquake.usgs.gov/fdsnws/event/1/query?eventid=ci37389218&f
ormat=geojson,0.008693,,85.0,"ci37389218",1.35,ml,,ci,26.0,"9km NE of Aguanga,
CA",0.19,28,"ci",automatic,1539475168010,"M 1.4 - 9km NE of Aguanga,
CA",0,earthquake,"geoserve,nearby-cities,origin,phase-data",-
480.0,1539475395144,https://earthquake.usgs.gov/earthquakes/eventpage/ci37389218
```

```
[42]: !awk -F',' '{print NF; exit}' earthquakes.csv # identify no of columns
```

26

```
[45]: headers = !head -n 1 earthquakes.csv
len(headers[0].split(','))
```

[45]: 26

```
[47]: df = pd.read_csv('earthquakes.csv')
df
```

```
[47]:
```

| | alert | cdi | code \ |
|------|-------|-----|------------|
| 0 | NaN | NaN | 37389218 |
| 1 | NaN | NaN | 37389202 |
| 2 | NaN | 4.4 | 37389194 |
| 3 | NaN | NaN | 37389186 |
| 4 | NaN | NaN | 73096941 |
| ... | ... | ... | ... |
| 9327 | NaN | NaN | 73086771 |
| 9328 | NaN | NaN | 38063967 |
| 9329 | NaN | NaN | 2018261000 |
| 9330 | NaN | NaN | 38063959 |
| 9331 | NaN | NaN | 38063935 |

| | detail | dmin | felt \ |
|------|---|----------|--------|
| 0 | https://earthquake.usgs.gov/fdsnws/event/1/que... | 0.008693 | NaN |
| 1 | https://earthquake.usgs.gov/fdsnws/event/1/que... | 0.020030 | NaN |
| 2 | https://earthquake.usgs.gov/fdsnws/event/1/que... | 0.021370 | 28.0 |
| 3 | https://earthquake.usgs.gov/fdsnws/event/1/que... | 0.026180 | NaN |
| 4 | https://earthquake.usgs.gov/fdsnws/event/1/que... | 0.077990 | NaN |
| ... | ... | ... | ... |
| 9327 | https://earthquake.usgs.gov/fdsnws/event/1/que... | 0.018060 | NaN |
| 9328 | https://earthquake.usgs.gov/fdsnws/event/1/que... | 0.030410 | NaN |
| 9329 | https://earthquake.usgs.gov/fdsnws/event/1/que... | 0.452600 | NaN |
| 9330 | https://earthquake.usgs.gov/fdsnws/event/1/que... | 0.018650 | NaN |
| 9331 | https://earthquake.usgs.gov/fdsnws/event/1/que... | 0.016980 | NaN |

| | gap | ids | mag | magType | ... | sources | status \ |
|------|-------|--------------|------|---------|-----|---------|-----------|
| 0 | 85.0 | ,ci37389218, | 1.35 | ml | ... | ,ci, | automatic |
| 1 | 79.0 | ,ci37389202, | 1.29 | ml | ... | ,ci, | automatic |
| 2 | 21.0 | ,ci37389194, | 3.42 | ml | ... | ,ci, | automatic |
| 3 | 39.0 | ,ci37389186, | 0.44 | ml | ... | ,ci, | automatic |
| 4 | 192.0 | ,nc73096941, | 2.16 | md | ... | ,nc, | automatic |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 9327 | 185.0 | ,nc73086771, | 0.62 | md | ... | ,nc, | reviewed |
| 9328 | 50.0 | ,ci38063967, | 1.00 | ml | ... | ,ci, | reviewed |

| | | | | | | | |
|------|-------|----------------|------|----|-----|------|----------|
| 9329 | 276.0 | ,pr2018261000, | 2.40 | md | ... | ,pr, | reviewed |
| 9330 | 61.0 | ,ci38063959, | 1.10 | ml | ... | ,ci, | reviewed |
| 9331 | 39.0 | ,ci38063935, | 0.66 | ml | ... | ,ci, | reviewed |

| | time | | title | tsunami | \ |
|------|---------------|--|--------------------------------------|---------|---|
| 0 | 1539475168010 | | M 1.4 - 9km NE of Aguanga, CA | 0 | |
| 1 | 1539475129610 | | M 1.3 - 9km NE of Aguanga, CA | 0 | |
| 2 | 1539475062610 | | M 3.4 - 8km NE of Aguanga, CA | 0 | |
| 3 | 1539474978070 | | M 0.4 - 9km NE of Aguanga, CA | 0 | |
| 4 | 1539474716050 | | M 2.2 - 10km NW of Avenal, CA | 0 | |
| ... | ... | | ... | ... | |
| 9327 | 1537230228060 | | M 0.6 - 9km ENE of Mammoth Lakes, CA | 0 | |
| 9328 | 1537230135130 | | M 1.0 - 3km W of Julian, CA | 0 | |
| 9329 | 1537229908180 | M 2.4 - 35km NNE of Hatillo, Puerto Rico | | 0 | |
| 9330 | 1537229545350 | | M 1.1 - 9km NE of Aguanga, CA | 0 | |
| 9331 | 1537228864470 | | M 0.7 - 9km NE of Aguanga, CA | 0 | |

| | type | | types | tz | \ |
|------|------------|---|-------|--------|---|
| 0 | earthquake | ,geoserve,nearby-cities,origin,phase-data, | | -480.0 | |
| 1 | earthquake | ,geoserve,nearby-cities,origin,phase-data, | | -480.0 | |
| 2 | earthquake | ,dyfi,focal-mechanism,geoserve,nearby-cities,o... | | -480.0 | |
| 3 | earthquake | ,geoserve,nearby-cities,origin,phase-data, | | -480.0 | |
| 4 | earthquake | ,geoserve,nearby-cities,origin,phase-data,scit... | | -480.0 | |
| ... | ... | | ... | ... | |
| 9327 | earthquake | ,geoserve,nearby-cities,origin,phase-data, | | -480.0 | |
| 9328 | earthquake | ,geoserve,nearby-cities,origin,phase-data,scit... | | -480.0 | |
| 9329 | earthquake | ,geoserve,origin,phase-data, | | -240.0 | |
| 9330 | earthquake | ,focal-mechanism,geoserve,nearby-cities,origin... | | -480.0 | |
| 9331 | earthquake | ,focal-mechanism,geoserve,nearby-cities,origin... | | -480.0 | |

| | updated | url |
|------|---------------|---|
| 0 | 1539475395144 | https://earthquake.usgs.gov/earthquakes/eventp... |
| 1 | 1539475253925 | https://earthquake.usgs.gov/earthquakes/eventp... |
| 2 | 1539536756176 | https://earthquake.usgs.gov/earthquakes/eventp... |
| 3 | 1539475196167 | https://earthquake.usgs.gov/earthquakes/eventp... |
| 4 | 1539477547926 | https://earthquake.usgs.gov/earthquakes/eventp... |
| ... | ... | ... |
| 9327 | 1537285598315 | https://earthquake.usgs.gov/earthquakes/eventp... |
| 9328 | 1537276800970 | https://earthquake.usgs.gov/earthquakes/eventp... |
| 9329 | 1537243777410 | https://earthquake.usgs.gov/earthquakes/eventp... |
| 9330 | 1537230211640 | https://earthquake.usgs.gov/earthquakes/eventp... |
| 9331 | 1537305830770 | https://earthquake.usgs.gov/earthquakes/eventp... |

[9332 rows x 26 columns]

```
[48]: df.to_csv('output.csv', index=False)
```



```
[55]: import sqlite3
with sqlite3.connect('quakes.db') as connection:
    pd.read_csv('tsunamis.csv').to_sql(
        'tsunamis', connection, index=False,
        if_exists='replace'
    )
```

```
[56]: with sqlite3.connect('quakes.db') as connection: tsunamis = pd.read_sql('SELECT_
    ↳ * FROM tsunamis', connection)
```

```
[58]: tsunamis.head()
```

```
[58]: alert      type      title \
0  None  earthquake  M 5.0 - 165km NNW of Flying Fish Cove, Christm...
1  green  earthquake      M 6.7 - 262km NW of Ozernovskiy, Russia
2  green  earthquake      M 5.6 - 128km SE of Kimbe, Papua New Guinea
3  green  earthquake      M 6.5 - 148km S of Severo-Kuril'sk, Russia
4  green  earthquake      M 6.2 - 94km SW of Kokopo, Papua New Guinea
```

| | place | magType | mag | time |
|---|---|---------|-----|---------------|
| 0 | 165km NNW of Flying Fish Cove, Christmas Island | mww | 5.0 | 1539459504090 |
| 1 | 262km NW of Ozernovskiy, Russia | mww | 6.7 | 1539429023560 |
| 2 | 128km SE of Kimbe, Papua New Guinea | mww | 5.6 | 1539312723620 |
| 3 | 148km S of Severo-Kuril'sk, Russia | mww | 6.5 | 1539213362130 |
| 4 | 94km SW of Kokopo, Papua New Guinea | mww | 6.2 | 1539208835130 |

```
[60]: import datetime as dt
import pandas as pd
import requests
```

```
[ ]:
```